

В основу процесса оптического распознавания текста (OCR) положены два основных принципа:

Сверточная многоуровневая нейросеть как биологический прототип OCR.

Ф.Уоссермен. Нейрокомпьютерная техника: Теория и практика.

F.Wasserman. Neural Computer Technics: A Theory and Practice.

Принципы оптимальной обработки сигнала.

Гоноровский И.С. Радиотехнические цепи и сигналы.

Gonorovskij Iosif Semenovich Radio Engineering Circuits and Signals.

## ОПИСАНИЕ ПРОГРАММЫ:

Механизм оптического распознавания текста реализован на C++, (основная среда разработки Xcode Mac OS) и портируется на другие платформы. Программа ориентирована на потоковое распознавание восточных текстов.

Печатные тексты и манускрипты на восточных языках используют алфавиты из большого количества букв, содержит вертикальные и горизонтальные лигатуры и обладают сложной и иногда не известной (утраченной) грамматикой. В восточных текстах речь идет об алфавитах в сотни и тысячи букв (в обычных алфавитах десятки букв). Лигатуры могут объединять несколько букв в вертикальные и горизонтальные стеки. Такие лигатуры могут занимать место в соседней строке, изменять форму составляющих их букв. С точки зрения распознавания это эквивалентно увеличению алфавита до нескольких тысяч букв.

Программа OCR разбита на несколько логических блоков:

1. Обработка битового изображения.
2. Оптическое распознавание текста на основе графической сверточной нейросети.
3. Классификация графических признаков и применение шрифтовой геометрии
4. Грамматический и словарный контроль на основе морфологической и статистической модели.
5. Интерфейс и управление клиент-сервер многопроцессорного распознавания текста.

Теоретическое описание алгоритмов работы логических блоков

### 1. Обработка битового изображения.

Обработка битовой картинки выполняется по известным принципам, такими как:

адаптивный порог, предварительная сегментация текста, определение наклона текста и исправление наклона текста. Нормализация текста включает в себя нормализацию толщины штриха шрифта на основе выделения фокальных линий, устранение пиксельного шума, выбор параметров оптимального сглаживания.

В результате получается одно битовая черно – белая нормализованная картинка в формате `csitt` `Fax 4` текста, готовая к распознаванию.

### 2. Оптическое распознавание текста.

Биологическим прототипом OCR является **сверточная многоуровневая нейросеть**, структура которой основывается на строении зрительной коры головного мозга. С аппаратной точки зрения зрительная кора головного мозга представляет собой структуру из элементов (разные типы нейронов) обладающих относительно небольшим быстродействием и очень высокой степенью параллелизма. Потреблением энергии всей зрительной корой головного мозга оценивается всего в несколько ватт. Современные вычислительные структуры наоборот обладают, по сравнению со зрительной корой головного мозга, не высокой степенью параллелизма (все процессоры в основном выполняют вычисления последовательно), высоким электрическим потреблением энергии и очень высоким быстродействием. Вычислительная мощность современных процессоров достигла уровня, после которого можно говорить о моделировании нейросети, структура которой основана на строении

зрительной коры. Модель элемента этой сети – неокогнитрон – показала высокую эффективность применения для распознавания образов.

Алгоритмы обработки, которые используются зрительной корой головного мозга при обработки изображений беспрецедентны по эффективности, гибкости, компактности и красоте. В современной научной среде накоплен огромный экспериментальный материал по функционированию зрительной коры головного мозга. Поэтому центральной задачей создания программы оптического распознавания текста был выбран метод моделирования параллельных алгоритмов реализованных в зрительной коре головного мозга, в структурах последовательного типа, т.е. современных процессорах. Эффективным подходом является сохранение идеологии параллельных алгоритмов, реализованных в зрительной коре головного мозга, при их реализации в структурах последовательного типа. При этом приходится отказываться от достижения внешнего биологического правдоподобия и обращая внимание только на достижение максимальной точности и быстродействия.

Как показывает практика, формальное моделирование алгоритмов параллельных структур в структурах последовательного типа, не эффективно (для прямой реализации и обучения нейросети быстродействие – серьезная проблема).

Поэтому практическая реализация ОСР в коде с++ совсем не похожа на нейронную сеть, в то же время максимально сохраняя идеологию параллельных алгоритмов нейросети.

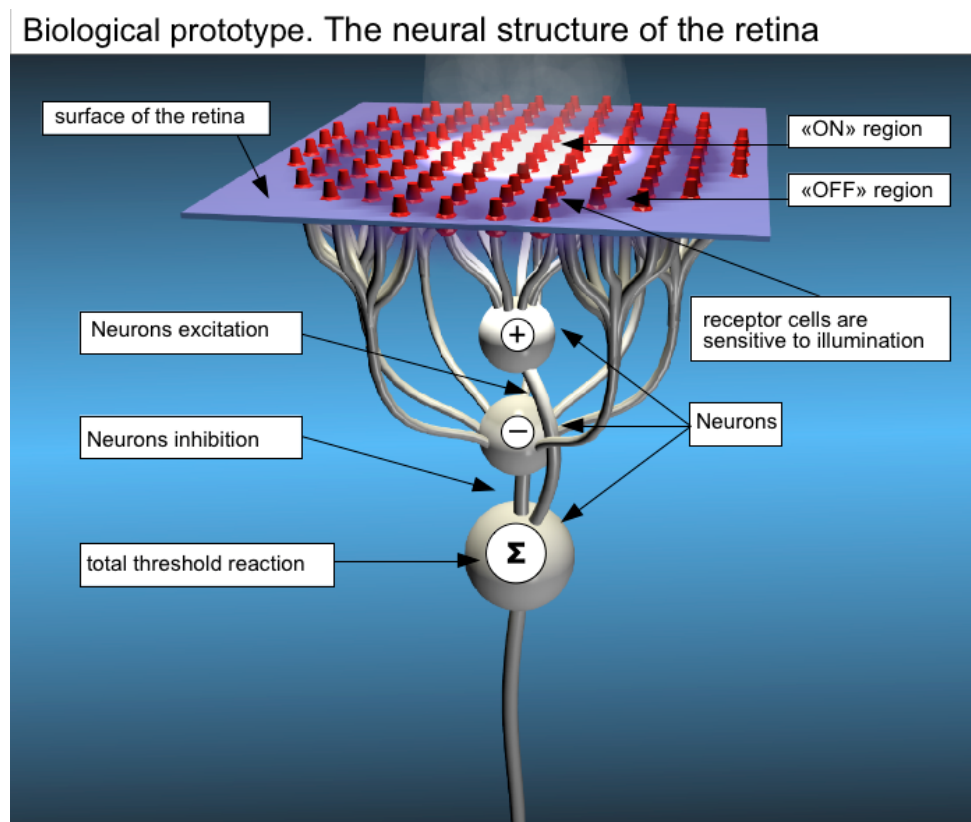
Реализованный подход можно описать в терминах кластерной сверточной нейросети. Первый слой нейросети в данной реализации это неокогнитрон размером от 12x12 до 32x64 пиксела, который используется для выполнения графической свертки по отношению к выделенным признакам изображения – ядрам свертки.

На втором слое выполняется градиентный поиск максимумов найденных признаков.

На третьем слое сети выполняется принятие решения по классификации признаков на основе обучающей выборки букв алфавита и сборки букв на основе правил шрифтовой грамматики.

Четвертый слой оценивает правдоподобность и корректирует гипотезы букв по отношению к словарю текстов и грамматическим правилам.

#### Описание биологического прототипа "движка" ОСР.



**Ris.1** Биологический прототип ОСР.

В основе алгоритма распознавания лежит базовый биологический принцип – принцип работы областей **ON** – **OFF**, показанный на рис.1. На поверхности сетчатки глаза сформированы две круглые области состоящие из светочувствительных колбочек (на рисунке красного цвета). Первая область это **ON** (на рисунке белого цвета) все колбочки попадающие в эту область

подключены к нейрону, осуществляющему возбуждение (+). Все колбочки попадающие в область **ОН**, концентрическая окружность описанная вокруг области **ОН** (на рисунке темного цвета), подключены к другому нейрону, осуществляющему торможение (-). Выходы возбуждающих (+) и тормозящих нейронов (-) подключены к суммирующему нейрону (с нарисованным математическим знаком сумма), который выдает пороговую реакцию пропорциональную разности освещённости и площадей центральной **ОН** и периферийной **ОФФ**, областей. Площади центральной **ОН** и периферийной **ОФФ**, областей равны.

Например:

Если осветить только центральную область **ОН**, то выходная реакция суммирующего нейрона будет максимальной положительной.

Если осветить только периферийную область **ОФФ**, то выходная реакция суммирующего нейрона будет максимальной отрицательной.

Если равномерно одновременно осветить и центральную **ОН** и периферийную **ОФФ**, области, то реакция центральной **ОН** и периферийной **ОФФ** областей будет одинакова и выходная реакция суммирующего нейрона пропорциональна разности освещённости областей, будет близка к нулю.

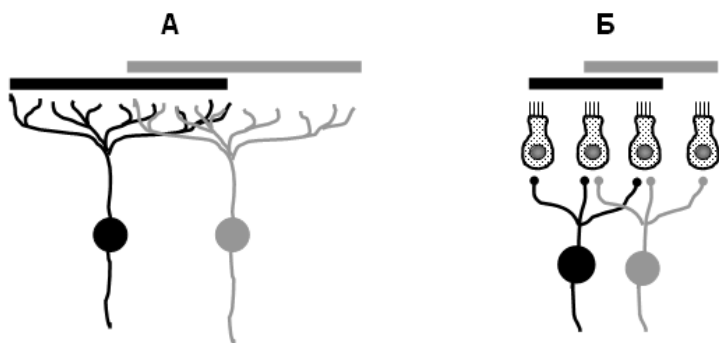
Если осветить центральную область **ОН** световым пятном размером меньше площади области **ОН**, то выходная реакция суммирующего нейрона будет не большой положительной и пропорциональна освещенной площади.

Если осветить центральную область **ОН** световым пятном размером больше чем площади области **ОН** с захватом области **ОФФ**, то выходная реакция суммирующего нейрона будет не большой положительной и пропорциональна разности освещенных площадей областей **ОН** и **ОФФ**.

Суммирующей нейрон выдаст максимальную положительную реакцию при условии **ОСВЕЩЁННОСТИ** центральной области **ОН**, одновременной **НЕ ОСВЕЩЁННОСТИ** периферийной области **ОФФ** и **БЕЗРАЗЛИЧЕН** к какой бы то ни было освещённости вне областей **ОН** и **ОФФ**. Для максимальной реакции важно что бы размер светового пятна точно совпадал с размером центральной **ОН** области и её положением, а яркость была постоянной по всей площади области **ОН**.

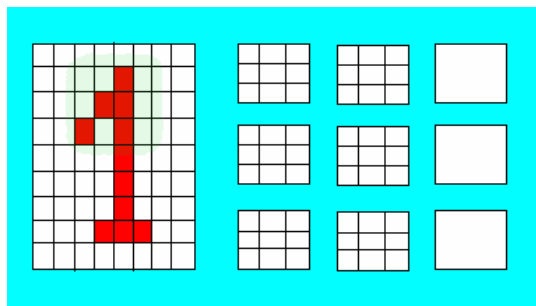
По сути дела подобная структура **ОН ОФФ** является детектором или фильтром, дающего максимальную реакцию на круглое световое пятна равномерной освещённости.

Существуют суммирующие нейроны с максимальной реакцией не только на световое пятна, но и на световые линии, границы освещенности и другие более сложные структуры.



**Ris.2** Структура расположения нейронов в сетчатке.

Следует отметить что на поверхности сетчатки глаза сформировано огромное количество областей **ОН ОФФ** которые взаимно перекрываются (их центры смещены на несколько колбочек), чем достигается их инвариантность к положению светового пятна рис.2 .



**Ris.3** Изображение процесса распознавания сверточной нейронной сети цифры «1»

<http://www.ampersant.ru/glaz/>

Предположим что ON и OFF области имеют не круглую, а более сложную форму, например соответствующей верхней части цифры единица "1" назовем её "признак" рис.3.

На рис.4 показана реакция признака соответствующая верхней части цифры "1" на весь распознаваемый текст. В правом верхнем углу показан признак. Полутонотом показан сам распознаваемый текст. Реакция признака просчитана для каждой точки распознаваемого текста. Синим цветом показана реакция признака на распознаваемый текст. Чем сильнее реакция, тем темнее синий цвет. Чем слабее реакция, тем светлее синий цвет. Видно что самая сильная реакция наблюдается в момент совпадения признака с верхней частью цифры единица "1". При достижении реакции порогового значения принимается решение о совпадении признака с верхней части цифры "1".

Признаков в одной цифре (букве) может быть несколько, например соответствующих верхней, средней и нижней части "1". Признак это часть цифры вырезанная из неё маленьким прямоугольником. Признаки могут соприкасаться или перекрываться. Совокупность нескольких признаков однозначно определяет цифру (букву). На рис.3 в виде анимации схематично показан процесс распознавания цифры "1" по трем признакам. Как видно из рисунка на каждом нейронном слое реализует разные уровни обобщения вплоть до принятия решения по цифре, входной слой чувствителен к простым образам, таким как признаки, в то время как, реакция следующего слоя является более сложной и не зависимой (например от положения образа), в нашем случае реакцией будет цифра "1". Инвариантность к положению признака также достигается применением большого количества областей ON OFF (форма которых соответствует форме признака), которые взаимно перекрываются ( т.е. их центры смещены на несколько колбочек).

Тогда структура ON OFF будет являться фильтром (детектором) близким к оптимальному(наилучшему) для одного из признаков цифры "1".

### Синтез фильтра, оптимального по отношению к заданному сигналу.

Существуют несколько фундаментальных положений связанных с этой задачей:

- "Задача синтеза фильтра, оптимального по отношению к заданному сигналу, действующего на фоне помехи с заданными характеристиками, решена в работах Н.Винера, А.Н.Колмогорова, В.А.Котельникова. Для задач обнаружения сигнала в шумах наибольшее распространение получил критерий максимума отношения сигнал-помеха на выходе фильтра. Фильтры отвечающие этому критерию, называются согласованными (оптимальными)". стр 442.

- "Сигнал на выходе согласованного фильтра с точностью до постоянного коэффициента А совпадает с корреляционной функцией входного сигнала.

$S_{вых}(t) = A B_s(t - t_0)$ .  $S_{out}(t) = A B_s(t - t_0)$ .

Где:

$S_{вых}(t)$  - сигнал на выходе согласованного фильтра.

$B_s(t - t_0)$  - корреляционной функции входного сигнала.

A - постоянный коэффициент.

t - время.

$t_0$  - время задержки."

Глава: Согласованная фильтрация сигнала на фоне помех. Стр 450.

Задачу оптимальной фильтрации сигнала на фоне помехи можно свести к вычислению корреляционной функции входного сигнала.

Другими словами, для реализации согласованного с сигналом фильтра, максимизирующего отношение сигнал-помеха на выходе фильтра, достаточно вычислить корреляционную функцию входного сигнала.

"Корреляционная функция входного сигнала:

$B_s(t_q) = [-\text{бесконечность} \int + \text{бесконечность}] S(t) S(t + t_q) dt$ .

Где:

$B_s(t_q)$  - корреляционной функции входного сигнала.

$S(t)$  - сигнал.

$S(t + t_q)$  - копия сигнала, сдвинутая на величину  $t_q$  по оси времени.

Отсюда видно, что  $B_s(t_q)$  характеризует степень связи (корреляции) сигнала  $S(t)$  со своей копией, сдвинутой на величину  $t_q$  по оси времени. Ясно что функция  $B_s(t_q)$  достигает максимума при  $t_q=0$ , так как любой сигнал полностью коррелирован с самим собой. При этом  $B_s(0)=[-\text{бесконечность} \int +\text{бесконечность}]S(t)S(t)dt$ , максимальное значение корреляционной функции равно энергии сигнала.

С увеличением  $t_q$  функция  $B_s(t_q)$  убывает (не обязательно монотонно) и при относительном сдвиге сигналов  $S(t)$  и  $S(t+t_q)$  на величину, превышающую длительность сигнала, обращается в нуль."

стр 79-80.

В качестве копии сигнала  $S(t+t_q)$  может выступать другой сигнал совпадающий по форме с входным сигналом. Этот сигнал можно назвать эталоном.

В случае OCR временную корреляционную функцию можно обобщить до пространственной двухмерной корреляционной функции. Таким образом, задачу пространственной двухмерной оптимальной фильтрации сигнала на фоне помехи, можно свести к вычислению пространственной двухмерной корреляционной функции входного сигнала. Корреляционная функция вычисляется между распознаваемым текстом и частью буквы эталона, называемой признаком.

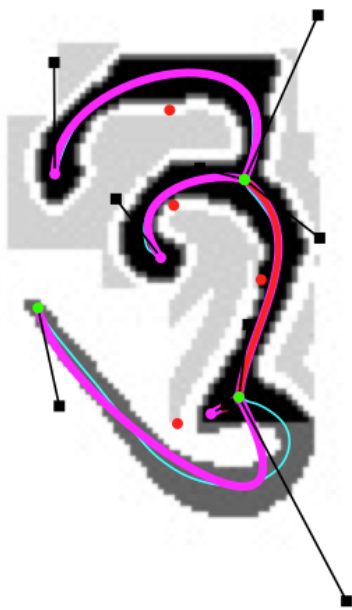
Максимальное значение корреляционной функции распознаваемого текста и признака буквы эталона называется коэффициентом корреляции и равно максимальной пространственной энергии признака буквы. Для получения максимальной гибкости и точности распознавания реально используются не целые буквы, а их части называемые признаками. Для получения максимального значения корреляционной функции используются градиентный поиск.

Возвращаясь к биологическому прототипу:

На поверхности сетчатки глаза сформировано огромное количество областей ON OFF которые взаимно перекрываются (их центры смещены на несколько колбочек), которые вычисляют пространственную двухмерную дискретную корреляционную функцию. Шаг дискретизации составляет несколько колбочек. Область ON так же суммирует с единичными весами (световое пятно обладает равномерной освещённостью) все что попадает в его площадь. Подобная структура ON OFF является пространственным фильтром близким к ОПТИМАЛЬНОМУ, для круглого светового пятна равномерной освещённости, размером в область ON и инвариантна к положению светового пятна. Как уже говорилось выше, существуют области ON OFF с максимальной реакцией не только на световое пятна, но и на более сложные формы: световые линии, границы освещенности и т.д.

### **Практическая реализация алгоритмов оск.**

На вход блока OCR подаётся одно битовая черно - белая нормализованная картинка текста (формат CCITT Fax 4) и выборка из базы данных состоящая из букв - эталонов, соответствующих алфавиту распознавания (формат XML). Буквы базы данных предварительно размечены на признаки. Каждой букве соответствует несколько признаков (верхняя часть рис.5, тонкие зелёные прямоугольники). Выделение признаков осуществляется программно в соответствии с характеристиками букв и реализуется на этапе обучения алгоритма. Этот процесс выделения признаков, основан на выделении фокальных (особых) точек и фокальных линий для каждой буквы. Фокальные точки - это точки пересечения топологических элементов письма. Для выделения фокальных линий и фокальных точек OCRlib использует алгоритм скелетизации и кривые Безье Рис.4. Если воспользоваться биологической аналогией, то создание и подключение базы данных букв - эталонов, эквивалентно быстрому обучению нейросети для узнавания этих букв.



**Ris.4** Фокальные точки и фокальные линии тибетской буквы преобразованные в кривые Безье.



**Ris.5** Иллюстрация основного процесса распознавания (OCR-процесса).

Процесс распознавания конкретной буквы показан на анимированном gif рис.5 (движение тонкого зелёного прямоугольника слева на право, по нижней распознаваемой строке). В верхней части рисунка в жирном красном квадрате изображена буква эталон с несколькими расставленными признаками (шесть зеленых прямоугольников). На данном рисунке это одна Тибетская буква "ла". Нижняя часть рисунка это распознаваемая строка (на данном рисунке это одна Тибетская буква "са" другая "ла"). Черный цвет внутри зеленых прямоугольников это область ON серый цвет область OFF. В точном соответствии с областями ON OFF признаков каждой буквы, в базе данных сформированы маски ON OFF. На каждую букву создаётся несколько пар масок ON OFF, именно они и используются при распознавании.

Для распознавания конкретной буквы берется её первый признак с парой масок ON OFF и накладывается на черно - белое изображение букв в строке. Подсчитывается суммарное количество черных пикселей изображения строки попавших в маску ON, аналогично подсчитывается суммарное количество черных пикселей попавших в маску OFF. Затем вычисляется их разность и делится на количество пикселей маски ON (т.е. на площадь маски ON). Полученная цифра это мера совпадения признака эталона с реальной картинкой. Таким образом мерой совпадения, мерой корреляции признака эталона и изображения в строке, является отношение:

$$\text{Correlation} = 100 * (\text{Sum\_bit\_ON} - \text{Sum\_bit\_OFF}) / \text{SUM\_mask\_ON}.$$

Соотношение домножено на 100 что бы коэффициент корреляции изменялся от 0 до 100 процентов. Площадь маски ON, SUM\_mask\_ON по физическому смыслу, всегда не равна нулю.

Если  $(\text{Sum\_bit\_ON} - \text{Sum\_bit\_OFF}) < 0$  то эту разность заменяем нулём.

Реально для подсчета писклей попавших в маски используется быстрый метод основанный на команде процессора \_porcnt64

<https://msdn.microsoft.com/en-us/library/bb385231.aspx>

Которая подсчитывают все единичные биты в целом 64-битовом числе.

Эта формула математически совпадает с формулой реакции суммирующего нейрона Аддера на рис.1. Фактически вычисляется интеграл по изображению в пределах маски ON и отдельно в пределах маски OFF. Подобный интеграл по площади является пространственным фильтром близким к оптимальному (для данных простых форм признаков), т.е. наилучшим с точки зрения отношения сигнал шум.

Затем признак дискретно смещается на несколько пикселей вправо рис.5, вычисляется коэффициент корреляции для новой позиции признака и запоминается. Если в этом положении вычисленный коэффициент корреляции лучше, сдвиг повторяется в этом направлении. Если на этой позиции ранее вычислялся коэффициент корреляции, например для другой буквы, то запоминается наибольший из них. В биологической модели, это соответствует взаимному подтормаживанию нейронов и выделение нейрона с наиболее сильной реакцией (рассматриваются нейроны на которые поступает сигнал примерно с одной и той же площади сетчатки глаза).

"Дискретное смещение вправо продолжается до тех пор пока коэффициент корреляции не превысит 50 процентов ( $\text{Correlation} > 50$ ). Если это произошло, то смещение вправо прекращается. На рис.5 движущиеся тонкий зелёный прямоугольник превращается в красный при  $\text{Correlation} > 50$  (последняя буква в нижней строке). Затем берется второй признак буквы из базы данных вместе с координатами относительно первого сработавшего признака и вычисляется коэффициент корреляции, если он более 30 процентов то берется третий признак. Таким образом проверяются все признаки, в нашем случае все шесть. Если у каждого из всех шести признаков коэффициент корреляции больше 30 процентов ( $\text{Correlation} > 30$ ), то буква считается предварительно распознанной. На рис.5 этот процесс показан в виде анимации, тонкий зелёный прямоугольник превращается в красный при  $\text{Correlation} > 30$ . Если хотя бы у одного признака коэффициент корреляции менее 30 процентов, то буква считается предварительно не распознанной и продолжается дискретное движение первого признака вдоль строки т.е. продолжается дальнейший поиск буквы. Окончательное решение по распознаванию буквы принимается после того как подобным образом будут проверены все буквы базы данных (весь алфавит по всей строке). В итоге, в местах где сработал первый признак, имеется набор коэффициентов корреляции от разных букв, из них выбирается наибольшей и соответствующая ему буква считается окончательно распознанной и обводится жирным красным квадратом."

**Определение оптимума расположения признака**



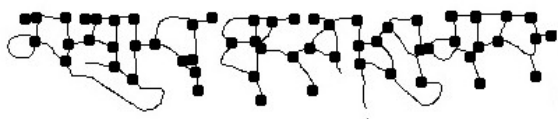


**Рис.6** Алгоритм поиска максимума корреляции признаков.

Для упрощения, ранее было опущено описание механизма самой юстировки признаков. В биологической модели, механизм самой юстировки соответствует постоянному движению глаза в небольшой области, движению подобному "тремору". На самом деле коэффициенты корреляции вычисляются не в точке, а в прямоугольной области сканирования (размером примерно один-два десятка пикселей), где и находится максимум коэффициента корреляции. Этот максимум считается вычислением "коэффициента корреляции в точке". Результат работы механизма сканирования показан на анимированном gif рис.6. Темно синим и серым цветом показаны маски признака, которым осуществляется сканирование. Черным буква в которой ищется данный признак. Красным область не совпадений. Зеленым область совпадений.

Механизм сканирования выполняется с применением известного алгоритма, который называется: "поиск максимума с помощью ЛЬВА в пустыне". ЛЕВ в пустыне работает поочередно по вертикали и горизонтали отыскивая максимум коэффициента корреляции в небольшой области размером примерно 8x9 пикселей. По сути дела это разновидность градиентного метода поиска максимума.

Для получения максимальной производительностью матричный алгоритм применяется только к буквам, фокальные точки которых удалось сопоставить с базой данных фокальных точек эталонов букв (Зеленые и пурпурные точки на Рис.4, черные точки на Рис.7). Применение алгоритма поиска фокальных точек перед матричным алгоритмом сокращает число циклов матричного алгоритма примерно на порядок.



**Fig.7** Фокальные точки и фокальные линии из текстовой строки тибетского манускрипта.

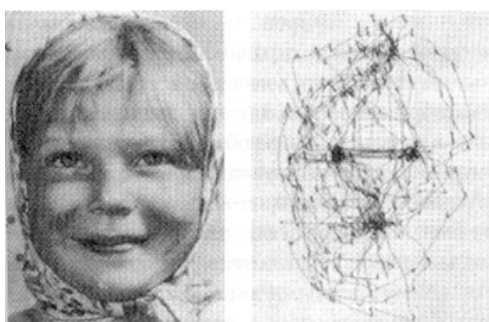
### **Деление буквы на признаки.**

"Необходимость деления буквы на признаки связана с тем, что реальные изображения букв прежде, чем попасть на распознавание, претерпевают различные тонкие деформации, причем их характер и величина могут изменяться даже в пределах одной страницы текста. Речь идет о тонких аффинных деформациях (изменениях: масштаба, поворота, сдвига и т.д.), нестабильности параметров растискивания при печати и просто случайных изменениях формы. Алгоритмы обработки, которые используются зрительной корой головного мозга, инвариантны к подобным тонким деформациям поэтому глаз их просто не замечает. Однако стоит перейти к



формальным алгоритмам распознавания текста, например применить механизм масок ON и OFF целиком ко всей букве, то проблема тонких деформаций встанет в полный рост. Коэффициенты корреляции между одинаковыми буквами становятся низкими и вся OCR не эффективна. Деление буквы на небольшие признаки (по сравнению с буквой), позволяет сделать процесс OCR инвариантным(независимым) ко многим видам тонких деформаций, встречающимися на практике. Реально тонкие деформации обрабатываются в пределах примерно 10 процентов от размера буквы.

Как было показано выше, при вычислении коэффициента корреляции признак ищет максимум коэффициента корреляции в небольшой области. Следовательно каждый признак обладает подвижностью в некоторой области (примерно 10 процентов от размера буквы). Поэтому всю букву можно представить как некий виртуальный векторный каркас, с признаками на концах векторов рис.7. Вектора подвижны и эластичны в механическом смысле. При юстировке признаки буквы эталона находят признаки в распознаваемой букве, каждый в своей небольшой области. Такая виртуальная механическая конструкция способна отработать практически все виды тонких деформаций, встречающимися на практике. В пределах признака тонкие деформации не велики и обрабатываются самым способом вычисления коэффициента корреляции."



Ил. 1. Слева фотография, рассматривавшаяся в течение нескольких минут. Справа схема движения взгляда по ней: скачки от одного пункта фиксации к другому прочерчены линиями (по Ярбусу).

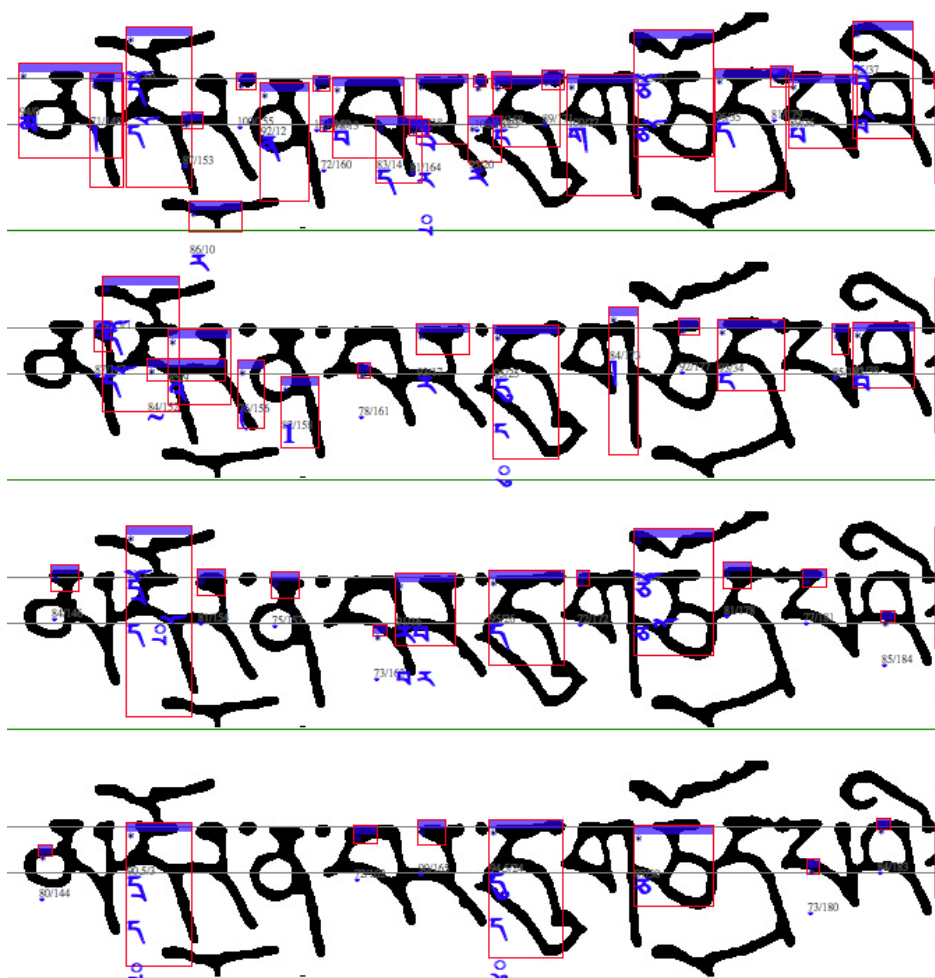
**Рис.8** Процесс рассматривания черт лица глазом человека в течении нескольких минут. Вы можете видеть фиксацию глаз в фокальных точках изображения.

(Ярбус Альфред Лукойнович [https://ru.wikipedia.org/wiki/Ярбус,\\_Альфред\\_Лукьянович](https://ru.wikipedia.org/wiki/Ярбус,_Альфред_Лукьянович))

Биологическим прототипом деления буквы на небольшие признаки является механизм сканирования глазом картинки, рис.8. На рисунке видно как при сканировании глаз выделяет особые точки (в нашей аналогии признаки) и внимательно их сканирует в небольшой области, а затем перепрыгивает на другой признак и т.д. Наиболее темные на рисунке области (почти точки) это признаки.

### 3. Грамматический и словарный контроль.

Гипотезы букв после завершения графического этапа распознавание с помощью сверточных нейронных сетей, являются основой для работы алгоритма вероятностных цепей Маркова. В качестве реализации этого алгоритма использовался полнотекстовый поисковый алгоритм Малькольма К. Харрисона [САСМ 14 (1971) 777-779 ], основанного на словаре (не четкий поиск). Словарь создается на основе всего доступного тибетского корпуса текстов. На предварительном этапе все базовые буквы объединяются в стеки в соответствии с правилами шрифтовой грамматики. Для каждой буквы на этом этапе мы имеем несколько гипотез с различной корреляцией.



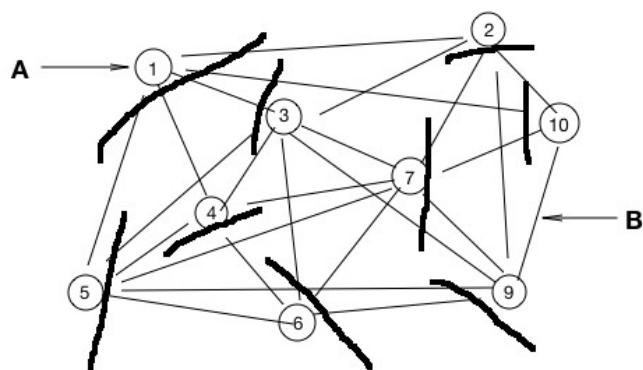
**Ris.9** Рукописные буквы в OCR ранжированные по корреляции. Верхняя строка имеет лучшую корреляцию.

На первом этапе алгоритма словаря все буквы гипотезы в текстовой строке объединяются попарно. На следующем этапе пары букв объединяются в слова и фразы. На этом этапе мы можем ожидать около 200 словарных гипотез на одну текстовую строку. Сборка фразы из словарных гипотез основана на графической корреляции каждой буквы гипотезы, частоте повторения и длине гипотезы. Распознанный текст окрашивается разными цветами. Синий цвет – точно распознанный текст. Красный цвет означает маловероятный текст. Зеленый цвет – промежуточная ситуация между двумя описанными выше.

Для распознавания рукописного текста необходимо решить задачу выявления паттернов и их сочетаний в графическом узоре. Нужно научить программу разделять задачу на этапы, выявлять закономерности, создавать на их основе гипотезы и подтверждать или опровергать эти гипотезы на основе опыта.

В первом приближении задачу можно рассмотреть как выбор метода графических сверток и задачу оптимизации Беллмана решаемую методом сверточной нейросети.

#### Весовой граф отображающий вероятности сочетания признаков букв



A. Выделенные признаки букв имеющие порядковый номер и вероятность с которой они принадлежат к классу признаков в классификации

B. Ребра графа, отображающие вероятность сочетания признаков

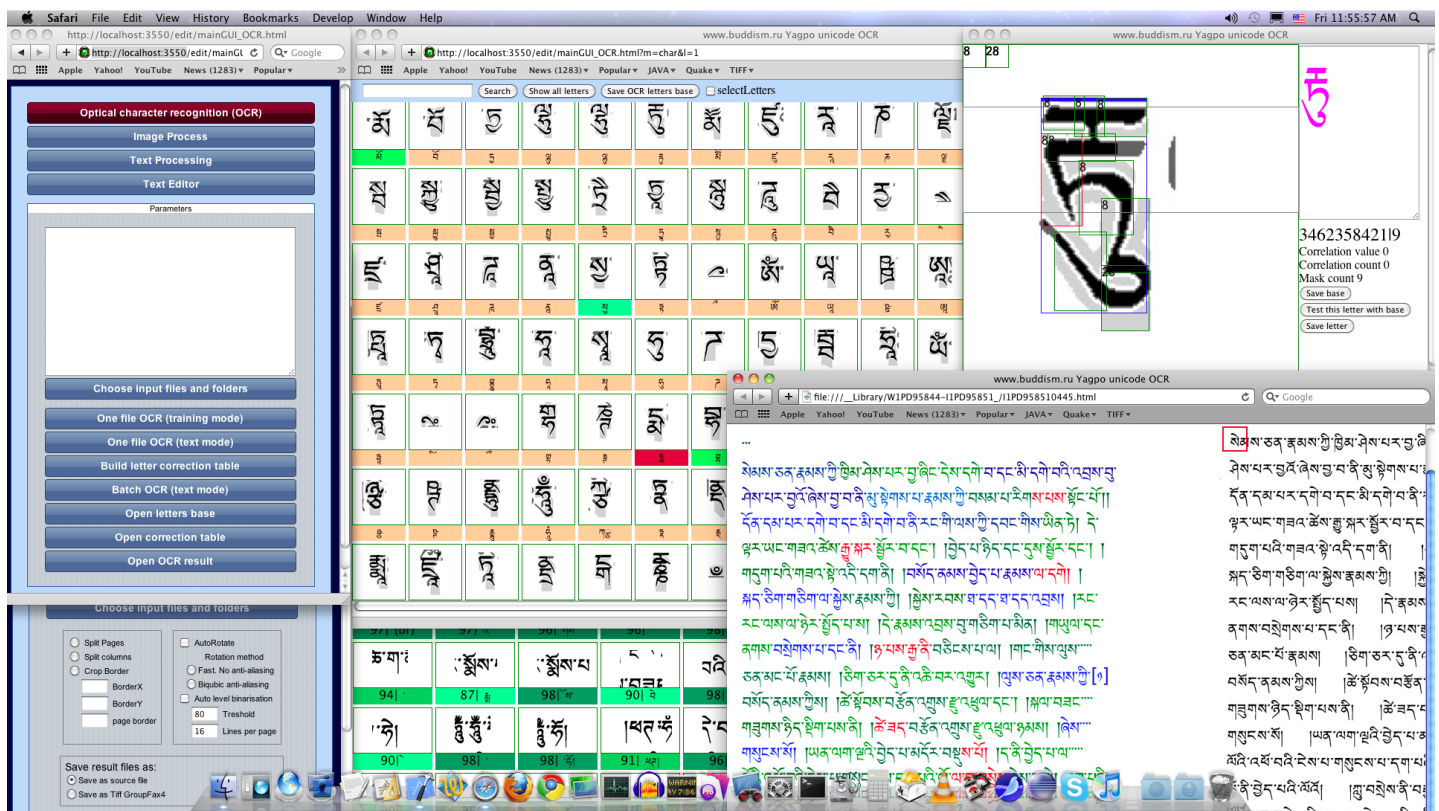
Однако сложность в том, что в отличие от печатного текста не удастся определить вероятность сочетания признаков на основе только размеченной обучающей выборки. Необходимо понять, как применить машинное обучение для поиска и классификации закономерностей распределения признаков на основе самого текста.

Решение этой задачи имеет применение и в других задачах машинного обучения. Например при выявлении графических образов, колоризации изображения и машинном переводе.

Одним из вариантов обучения является программное создание онтологии. Этот подход применяется при машинном переводе.

#### 5. HTML интерфейс и коммуникация клиент-сервер в аспекте потокового распознавания текста.

Процесс распознавания каждой странице является независимым, обмен с базой данных осуществляется через mpar. Это дает возможность использовать OCRLib как эффективный OCR-сервер. Каждому процессу требуется около 400 МБ памяти, распознавание страницы в один поток выполняется за 1,5 минуты. На 12-ядерном сервере с 24Gb памяти можно эффективно обрабатывать 24 страницы одновременно.



**Ris.10 OCRLib** общий пользовательский интерфейс программы.

Интерфейс программы выполнен на базе HTML рис.10 (левый верхний угол) и архитектуры клиент-сервер на основе <https://embedthis.com/esp/>. GUI программы основан на HTML5 и может использоваться с любым браузером, например Safari, Chrome и т.д. или пользовательский интерфейс программы. Это позволяет дистанционно управлять процессом сетевого распределённого оптического распознавания текста. Данные, необходимые для работы программы хранятся в базе данных состоящая из букв – эталонов, соответствующих алфавиту распознавания (формат **xmL**), на рис.10 желтая таблица в центре картинки. База данных также обеспечивает доступ к корпусу тибетских текстов, поиск в библиотечном каталоге, а так же по словное сопоставление тибетского и английского текста, тибетского и русский текста. Распознанный текст выдается в формате **HTML,RTF** он показан в правом нижнем углу. Черный это исходный распознаваемый текст, небольшая красная рамка удобна для синхронизации букв распознанного и исходного текста. Распознанный текста размечен разными цветами. Синий цвет – точно распознанный текст. Красный цвет – цвет вероятных ошибок, если ошибки в тексте есть, то их следует искать в красном распознанным тексте. Красный цвет это не ошибки. Зелёный цвет – промежуточная ситуация между двумя описанными выше (практического значения не имеет, слишком мала вероятность ошибок). Программа также имеет базовый графический интерфейс Cocoa для MacOS.

#### ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ:

В настоящее время на базе OCRLib функционирует бесплатный онлайн-сервис OCR [www.buddism.ru/ocr](http://www.buddism.ru/ocr). Эта работа выполнена в рамках проекта буддийской Библиотеки с открытым исходным кодом. В течение 2010-2017 годов было обработано более 500 000 страниц для TBRC.org, Trase Foundation, монастыря Сера, Государственного университета Санкт-Петербурга и многих других отдельных исследователей и организаций. В настоящее время с помощью сингапурской сангхи выполнен первый этап разработки алгоритма распознавания тибетских рукописей с точностью распознавания текста 85%.

Пример распознавания буддийского корпуса текстов Кангюр:

[http://www.buddism.ru///CANON/ KANGYUR MAIN/bka' 'gyur Beijing OCR tiff/?zip=bka%27 %27gyur Beijing OCR tiff v064.zip&img\\_page=bka%27 %27gyur Beijing OCR tiff v064%2Fpage\\_0436.tif&ocr=1](http://www.buddism.ru///CANON/ KANGYUR MAIN/bka' 'gyur Beijing OCR tiff/?zip=bka%27 %27gyur Beijing OCR tiff v064.zip&img_page=bka%27 %27gyur Beijing OCR tiff v064%2Fpage_0436.tif&ocr=1)  
[Http://www.buddism.ru///CANON/ KANGYUR MAIN/bka' 'gyur Beijing OCR tiff/](http://www.buddism.ru///CANON/ KANGYUR MAIN/bka' 'gyur Beijing OCR tiff/)

## ПЕРСПЕКТИВЫ:

1. Дальнейшее увеличение точности распознавания и сведение необходимости корректуры распознанных текстов к нулю.
2. Дальнейшее увеличение гибкости и масштабо независимости распознавания.
3. Увеличение быстродействия.
4. Увеличение скорости создания баз данных признаков для новых языков.
5. Распознавание рукописного текста.

## ССЫЛКИ:

Репозиторий кода:

<https://github.com/RimeOCRLIB/OCRLib>

GUI, документация и репозиторий баз данных:

[http://www.buddism.ru///ocrlib/SOURCE\\_CODE/](http://www.buddism.ru///ocrlib/SOURCE_CODE/)

Почта для связи:

Gomde@mail.ru

## РИСУНКИ:

1. 1\_OCR\_biological\_prototype.png Биологический прототип OCR.
2. fig\_05\_5.gif Структура расположения нейронов в сетчатке.
3. 2\_recon\_show.gif Картинка с анимированным gif "1"  
<http://www.ampersant.ru/glaz/>
4. 5\_OCR\_virtual\_vector.jpg Виртуальный векторный каркас, с признаками на концах векторов.
5. 3\_OCR\_letter.gif Анимированный gif с расставленными признаками - прямоугольниками и процессом OCR.
6. 4\_OCR\_letter.gif Анимированный gif с демонстрацией механизма самоустойки признаков.
7. FocalPoints.jpg Фокальные точки и фокальные линии из текстовой строки тибетского манускрипта.
8. 6\_OCR\_eye\_mov.jpg Движение глаз при рассматривании изображения (по А.Л. Ярбусу): [http://www.humanbrain.ru/luria/luria-2\\_03\\_c01.htm](http://www.humanbrain.ru/luria/luria-2_03_c01.htm)
9. letterOCR.png Рукописные буквы в OCR ранжированные по корреляции. Верхняя строка имеет лучшую корреляцию.
10. 7\_OCR\_Lib.png Снимок с HTML интерфейсом, базой данных, распознанным и исходным текстом, корректурными таблицами.

## ЛИТЕРАТУРА:

Гоноровский И. С. Радиотехнические цепи и сигналы. Сов. Радио 1977.  
Gonorovskij Iosif Semenovich. Radio Engineering Circuits and Signals.  
Soviet Radio 1977.  
[http://publ.lib.ru/ARCHIVES/G/GONOROVSKIY\\_Iosif\\_Semenovich/](http://publ.lib.ru/ARCHIVES/G/GONOROVSKIY_Iosif_Semenovich/)

Ф. Уоссермен. Нейрокомпьютерная техника: Теория и практика.

<http://www.ampersant.ru/nntech.pdf>

F. Wasserman. Neural Computer Technics: A Theory and Practice.

Дональд Эрвин Кнут. Искусство программирования.

Donald Ervin Knuth. The Art of Computer Programming.

Генри С. Уоррен (младший). Алгоритмические трюки для программистов.

Henry S. Warren, Jr. Hacker's Delight-2.

Вилли К, Детье В, Биология.

Willy K, Dete V, Biology.

[http://publ.lib.ru/ARCHIVES/V/VILLI\\_K/\\_Villi\\_K..html](http://publ.lib.ru/ARCHIVES/V/VILLI_K/_Villi_K..html)